

## VAST 2017 投稿修改计划

R2-primary

R4-secondary

### 一、Summary review 说了什么

综合打分是 2, reject.

Reviewers all think that this paper describes an interesting system, but there is a lack of clarity and sufficient information that makes this paper hard to accept at this point.

The following issues are brought up by multiple reviewers:

- The radar metaphor is only touched upon briefly, what is it good for and why is it in the title (R2, R3, R4) **Radar 隐喻在哪啊? 还在做当中**
- Some core points are unclear: whether streaming is really necessary/performed (R4), why are multiple trees shown (R1), why a decision tree? (R1) **tree 模型和实用价值, 和”一棵树”的说明方式对不对, 这棵树不断在变, 所以画出来了很多树.**
- Case study is not convincing (R2, R3, R4) **case 不够强**

各位 reviewer 的总结性评价(中重点语句)是:

R2: 这个系统有些描述的非常不清楚, 评估有点太局限了.

R4: 两大点没说好(**为什么选择这个模型啊?** 时间离散化的原因及种种带来的疑惑), 不能令人信服的 argue 这个系统的优势.| 写的看起来太 rush 了.

R1: 表达不清是主要问题. 此外, 尽管项目有实际的需求, 但是很多问题在系统并没有被 address.

R3: 尽管是个 design study, 却缺乏很多关于设计过程和评估的信息.

### 二、分类

按照惯例, 分为简单、中等、难三个不同难度.

## Easy

R2:

**4** I'd like to see a better walkthrough of the user interface. There is a lot of text in 6.3 and 6.4 that I find hard to connect to the figures. Some of the figures are very nice and clear (like Figure 4), but others are very poorly labeled (all the ones in the usage scenarios). This would help a lot to understand the system. **图片好好注释一下**

**6** The column break on the last page is unnecessary. There is no need to use the 10th page if there is nothing to add.

R4:

**4** The last section "Design Alternatives" in 6.3 is a little confusing. It starts by stating a radar screen metaphor was tested but discarded, then a little to tersely explains a two-level fisheye view (which would actually be a table lens if applied on the tabular Recommendation View), but then spends another full paragraph including a numbered formula and a figure on explaining the visual mapping of this (supposedly discarded) view. Please clarify.

**这里写太乱了 reviewer 完全理解错了**

**5** Section 6.4.1 Individual Panel description should be improved. From the description I assume time is mapped on the X axis with groups of bars per time step. Contrary to the text, Fig.1(D)

only shows four types of bars, not 5. If it indeed shows groups of 4(5) bars per time step, the display could be improved by introducing small gaps between time steps to ease perception of time step boundaries. Please clarify. In the following explanation of the detailed activity dynamics line chart, please add a forward reference to Fig.7 illustrating them. 这里 4 和 5 不对应 当时搞乱了 后来已经改了| 线图加引用标注

6 Incidentally, Fig.7 appears to only show dynamics lines for the red category (also, a legend is missing from either Fig.1 or Fig. 7). 只有红色为这一类动作; 此外所有图都没有 legend. Moreover, dot sizes seem to be uniform, at least in the printout I could not discern size differences contrary to the explanation in 6.4.1. An unlucky choice of the depicted data example? 可能已经来不及写了所以没有做好点大小这个编码

7 Section 6.4.2 also does not fully explain the group view, Fig.1(E). If time steps are per line and each cell has bot/human avatar activity counts, what is encoded per table column? The reference to Fig.1 is mislabeled as Fig.4. 这里的引用写错标号了

8 Section 6.4 is too brief, especially given the authors still have almost a full page to spare. The utility of the "red" and "yellow curves" does not become clear, or how they are used in interactive tree pruning. Please expand. Especially the discussion of interactive pruning and splitting operations could benefit from a video demonstration, consider adding one as complementary material. 两种线没有具体陈述

9 Section 7.2.1, explanation of Fig.6: contrary to the text, at time steps 6 and 9 in 6(A), p1 has the lowest frequency in the blue category, not the highest among players? Also, time steps are hard to identify due to lack of axis labeling. the reference to Fig.6(A) should be 6(B) (activity 3, presumably). Last paragraph: where in the displays player 8 is shown? Is this an accidentally truncated third use case (as 7.2.2. starts with "The third case..."). 这里也是写乱了

R1:

7 There are some places that the paper contradicts itself. For example, in Sec. 6.3, paragraph of "Design Alternatives", it says due to limited screen space, a sortable table representation was used, as opposed to a radar chart. But after several paragraphs, the radar view is introduced and it is also later used and mentioned in the case study (though there is no corresponding images shown in the case study). 这两个雷达图不一样, 澄清即可

8 The description and symbols used in algorithms are confusing. 公式算法美化

9 How many classes of avatar activities are categorized exactly? Some places say 4, while some say 5. 玩家的活动类的个数, 前后说的不一致了, 修改即可.

R3:

1 Section 1, para 2. It's mentioned that network-side methods lead to network overload, but it's not clearly stated why. Do you mean that it is computationally expensive to collect and analyse the network traffic? Or do the known data collection or analysis techniques somehow add to the existing data traffic? network-side methods. 传统方法劣势, 稍微补充下即可

5 Section 3.1, first para. You should explain the difference between avatars and players with some more detail. Also, it's mentioned that for this study, the focus is on avatars, but would the same be true for a deployed system where analysis is performed by operators on players? If not, and the focus will be on players, are there any issues that might impact your design? 澄清 avatar 和 player 之间的关系

6 Section 3.1. In describing the data, it's mentioned that timestamps are used. Are these periodic (e.g., every second) or event-based? 澄清下时间戳

7 Section 3.1. The following clear and concise sentence should be much earlier in the paper (e.g., on first read of the paper, it wasn't clear to me, till page 3, who the users of the system would be): "This work aims to design a visualization system, integrated with a machine learning model, to facilitate analysts from the game operator to find game bots based on the log data of avatars." 这句话提前, 比如直接给 intro, 告诉大家系统的目的, 放在 intro 最后? 可能是前面 intro 写的太过于学术化

12 Figure 4. How does this radar view link to what is shown in Fig. 1, e.g., how would I access the radar view from Fig. 1? 没讲清楚 radarview 怎么来的, 从哪里进去可以得到

## Medium

R2:

1 I don't find the algorithms terribly useful. They're mostly of the type: if new data, update model; if better model, replace current with better. I'd find it more useful to use that space to sketch the way the algorithm works with a few sentences that tell the reader the parts they actually need to know. Parsing the algorithms and figuring out how they relate to the visualization is quite tedious. 算法写太多了, 没意思 | 或者说清楚为什么写这么多。

2 Also, isn't line 4 in Algorithm 2 wrong? Why update the current tree if there's an alternate one and not the alternate? 算法错了? 并没有, 要解释下?

3 The radar view appears to be the least useful, and yet it's the most prominent in the title. It doesn't come up in the case studies. Did it not turn to be of value? I think it's fine if that's the case, but I feel like this needs to be spelled out. Radar 视图我们还在改进

R4:

2 Second, the timeline of avatar activities are prepossessed into time-stamped records of activity counts (frequencies). I assume this means discretization by some (small) time interval rather than arbitrary, high-resolution time instances. This implies that each decision trees only ever sees a subset of isolated snapshots of a real player's avatar or a bot's behavior. Is this sufficient to capture complex activity sequences, i.e., I image relevant/suspicious patterns may typically well span multiple time slots? 时间离散化问题, 信息丢失问题. 1) 网易那边也是这么类似的处理方式, 不知道能否拿来背书 2) 信息是否丢失需要再细考量, 因为本身我们就没有画出所有的树. 需要在这里作一个权衡。

R1:

1 (同 R4 2) It is not clear to me why there are multiple trees displayed in the dynamic tree view. My guess is that every user will form a tree, but in such case how the visualization method scales when there are potentially millions of users in the dataset? 模型没描述清楚以至于不知道为什么会有多棵树

2 Sec. 6 introduces the visualizations employed in the system. However, only few images are provided. With only textual description, it is hard to imagine what the corresponding visual stimuli was shown to the user to support the necessary analysis. It should be very critical in this work as it will help readers understand how the analysts' requirements can be fulfilled by each of the visualizations. 多加图文表述

3 The user requirements (Sec. 3.2) and the design guidelines (Sec. 6.1) are kind of repetitive and can be combined and reduced to create some space for other more important issues, such as provide images for explaining the visualizations. 用户需求和设计准则有重复的地方，可以删除一些东西

R3:

3 Section 2.3. I don't quite follow the conclusion in the final paragraph. I understand the reasoning for using a decision tree, based on your requirements, but I don't follow the connection to the point about the more complex correlations. 这一段讲动态结构可视化相关工作，收尾比较模糊，不够有力

4 Section 2. I found parts of this section not sufficiently clear in terms of justifying the critical work. In particular, several points seem to critique the previous work relative to requirements that aren't presented till Section 3. For example, the reasoning at the end of 2.1 (to allow analysts to "dig into" the process), and the need to take concept drift into account (2.2). It might be simple enough to fix this with a forward reference to the requirements at the start of Section 2, and even a quick, one-sentence high-level description of the requirements. 有些问题，挑战没在相关工作提到，一直到了 sec3 才提到，需要前置 | 这里需要考量，要不要这么做。

8 Section 3.1, item "Large volume of activities". There is mention of "140 types of activities". At what granularity are these? Might this classification of activities be different between operators? More importantly, it's not clear how this classification came to be from the "preliminary analysis of the data with analysts from the game operator". In particular, was the purpose of this preliminary analysis to provide a list of activity types for your tool (in which case, would such manual analysis have to be done by each operator), to tag different activities by avatars, or perhaps to do a cleansing of the data? It would be good to know more about this "preliminary analysis" in order to better understand the pre-conditions required for your tool. 对于 preliminary analysis 需要说的更为坦诚和明白一点。

9 Section 3.1, item "sparsity of data" and Section 3.2. It's mentioned that the percentage of bots is low. But how was this determined? Was it determined from the preliminary analysis done by the analysts, or was it a result of your later analysis with your model? If there was a count of the number of bots by a preliminary analysis, and later count by your tool, did these counts correspond? Was the preliminary analysis intended to obtain such a ground truth bot analysis in order to validate the effectiveness of your tool? 就是说，我们的数据是 label 好的，这个问题虽然前面提到，但是为了说明白的话，可能还是需要结合模型来说一下

13 Section 7. Why were these particular case studies chosen? Were they suggested as challenging in some way by the analysts? Without such a justification, one could suggest that you may have chosen cases that are advantageous to the strengths of your design.

14 Section 7.3, user feedback. Were the two analysts used for the evaluation different from the analysts used for the requirements gathering and design (they should be different)? 收集反馈和收集需求的人是同一批吗? 应该是一批人 or 两批人才比较合理. 这一点需要思考一下

15 Section 7.4, limitations. There are limitations of the proposed system, but I don't see limitations listed for your design and evaluation process. For example the evaluation was done with a limited number of analysts at one operator.

## Hard

R2:

5 The user feedback in 7.3 is the usual stuff you get from people being nice, but it's not very informative. The case studies are described in a way that's quite abstract and hard to get a good picture of how well the system really works. 用户反馈不够强 | 需要再去访谈

R4:

1 However, there is one critical issue where the exposition must be improved, with two sub-aspects: **First, the choice of the (only) modeling approach.** The authors state clearly that model interpretability has high priority which may indeed rule out other, more complex approaches like frequent sub-sequence mining or pattern mining approaches (see e.g. [A]). However, have other methods also been probed 探查 and discarded due to this necessary trade-off or limitations in streaming settings? Moreover, **why is support of the streaming setting a prerequisite in the first place?** Do game experts really need to analyze (and fine-tune!) these detection models in real-time? I somehow doubt that -- in short order, yes, but surely not as a game session is running. Thus, I am not at all convinced that the constraint of streaming data capability is inherent to the domain problem. From the paper, it more feels like a rather bothersome way to capture the temporal dynamics of player activities (see next).

对于模型的质疑\担忧, 选择这个模型的合理性.

3\* This critique(批评) is further substantiated by the information provided on the use cases/qualitative user evaluation, as well as mentioned during the discussion of limitations in Section 7.4. **One of the main cognitive tasks of the experts is to reason about per-time step changes between subsequent decision trees**, rather than the model and visualizations directly capturing and comparing the dynamics of avatar behavior across longer time spans. The users appraised the system, but it also stated they did not have any prior visualization experience. The case study therefore **fails to give convincing argument that the chosen visual analytics workflow**, which includes the choice of modeling technique, really is appropriate and efficient; and consequently, whether the qualitative result is more due to the baseline of the users not having access to any visual support prior to the study. Compare, for example, visual encoding proposed in paper [A] which is also on visual analysis of the dynamics of complex, long running temporal dynamics of player activities (albeit with a different focus of balancing game mechanics). **reviewer 认为我们做的东西让他们用了以后实际他们做的任务并不是我们想要的; 另外参与评估的人没有可视化经验也就不符合这个 workflow 中的要求.**

R1:

4 **The major limitation of the system is its scalability.** As also described in the last paragraph of Sec. 7.3, the proposed system would face scalability issues when there are larger numbers of avatars, classes, types of bots involved in the analysis. **可扩展性的问题是这个系统大问题**

5 It is mentioned in the paper that the system can be used as a diagnostic tool so that users can further optimize the bot detection model. However, after reading the paper it is still unclear to me how such a purpose can be served using the proposed system. **模型进一步 refine 可以实现? | 一定程度上个调优功能是否还必要, 有待商定, 后面可能会砍掉这个功能.**

6 (同 R4 1) There are several machine learning, data mining approaches designed for game bot detection, **why the decision tree model was selected over the others?** According to the related work, it seems decision tree is not the most effective model for such a purpose. Maybe because a decision tree is just easier to visualize and understand? **模型选择问题**

R3:

2 (同 R1 6 和 R4 1) Section 2.2. **I don't quite follow the conclusion that existing methods cannot be applied.** The end of the first paragraph seems to suggest that FVDT and CVFDT would be suitable since they were designed for stream data. **模型选择问题**

10 Section 3.2 (user requirements) and section 6 (visualisation design). I don't quite understand the process used here, or how the analysts from the game operator were involved. For example, 3.2 mentions that the "requirements are collected and refined iteratively", but it's not clear how this was done. If the analysts provided requirements on day 1, what process was involved in getting an iteration of the requirements? Did you work with the analysts individually, or as a group of three? For the "design guidelines" of 6.1, it's not entirely clear to me how these differ from your requirements. Were the analysts shown some initial designs and asked to annotate them, from which you iterated some new designs? Or were the analysts asked to brainstorm their own initial designs? For the specific interface items of sections 6.2, etc., it's not clear to me whether the analysts had an influence on the position of the control panel (for example). **需求和设计准则重复的比较多, 并且这个改进的过程写的不够真实. 需要进一步的细化的描述**

11 Section 6.3. Related to the above point, there is some mention of "Design Alternatives" here, which suggest the iteration of designs, but the overall description in the paper currently lacks information about this. **同上面这个问题**

### 三、感想

首先, 几位 Reviewer 评分是这样的. R2 是主审, 给了最高的 3 分. R4 是 secondary, 给了 2.5 分. 其余 R1、R3 都是 external 也都是 2.5 分. 其实 2.5 应该只是给点面子的分数, 最后的 summary 是 2 分, reject. 以文章完整性来说, 我深知其有多不完整, 以至于很多东西 reviewer 根本没抓到; 他们几乎每个人都在说我表达不清.

今年的 reviewer 没有特别激烈的语气, 这也可能和没写完有关, 他们无处可喷. scalability 这个问题也只提到了一次. 大多数的意见都是比较 trivial 的.

模型选择、对系统描述、和评估过程是主要问题. 修改主要就集中在这三大块.

由于我个人还是认为今年客观上时间不够，后面一再时间点脱节导致了比较惨的论文，从我个人角度来说经历比较非典型性，大部分 review 意见不能说对我今后投稿有较大的启发，所以不再描述。

初步的时间安排

**easy** 部分 10 天，由于论文本身不成熟，整个体系可能会有变化，前面的时间还要继续对一些可视化效果做改进。

**medium** 部分 + **hard** 部分，需要更复杂的思考以及作用户调研等，还需要迭代修改，时间不好预估，暂定 20 天。

制作视频等需要 3 天或更多时间，目标最晚在 7 月中旬投出去/